Performance Analysis of a Fully-Connected, Full-Duplex CDMA ALOHA Network with Channel Sensing and Collision Detection

Fook Loong Lo, *Member, IEEE*, Tung Sang Ng, *Senior Member, IEEE*, and Tony T. Yuk, *Member, IEEE*

***Abstract*— In cases where machines having bursty data are equally likely to transmit to one another, code-division multiple-access (CDMA) ALOHA which allows for an individual "virtual channel" for each receiving station may be a better multiple-access protocol than simple ALOHA. With the use of "receiver-based code" multiple-access protocol, it is also possible for a station to listen to the channel of the intended receiver before transmission, and also abort transmission when it detects others transmitting on the same channel. This paper describes a model for a fully-connected, full duplex, and slotted CDMA ALOHA network where channel sensing and collision detection are used. The model is analyzed using a discrete time Markov chain and some numerical results are presented. For a system with a large number of users, where Markov analysis is impractical, equilibrium point analysis is used to predict the stability of the system, and estimate the throughput as well as the delay performance of the system when it is stable. Finally, a comparison is made with a simple channel sense multiple-access with collision detection (CSMA-CD) network, showing that a substantial improvement in the performance is achieved by the proposed network.**

I. Introduction

SINCE the ALOHA wireless channel access protocol was proposed in [1], it has been shown to be a very simple and effective way of allowing many stations with infrequent short queries to communicate with a central computer. For networks where the machines are equally likely to communicate with one another, addresses of recipients can be attached to every message without changing the way each station access the common channel. With the use of the (CSMA) technique [2], or CSMA with collision detection (CSMA-CD) [3], the performance of a network can be quite good. This is amply demonstrated by the existence of the many local-area networks (LAN's) using Ethernet [4].

One drawback of networks which allow for addressing is the possibility of collisions between messages meant for different machines. Assigning different channels to the different machines using frequency-division multiple-access (FDMA) or time-division multiple-access (TDMA) is not a reasonable solution in cases where the machines do not transmit most of the time. The channels that are set aside remain unused and thus waste resources. Code-division multiple-access (CDMA) channels if unused, on the other hand, do not form such a waste. A lightly loaded spread-spectrum channel looks just like low-level noise to other narrow band users [5].

One way of creating channels using spread-spectrum signaling is by assigning different spreading sequences with which each station will receive messages. This "receiver-based code" protocol is investigated in [6], where the performance of a network of machines which can only transmit or receive at any one time is presented. The channels can also be created by using different phases of the same maximal length spreading sequence [7], [8] if the system is synchronized.

In this paper, a more sophisticated version of a CDMA ALOHA LAN is presented and then analyzed. Each receiver is still assigned a "virtual channel" with which it receives messages, but now each sender will listen to the channel of the intended receiver before transmitting. The stations will also listen to the channels during transmission and compare the received signal with the transmitted signals. Should collisions be detected, the transmitting stations will cease transmission and try again later. This access protocol is therefore equivalent to that of p-persistent CSMA-CD of simple ALOHA as described in [2].

It must be pointed out that there are many technical challenges involved in implementing this system. If different spreading codes are used, the system must be robust enough so that even with degradation in reception caused by different signal strengths (near–far problem) and multipath fading, an individual receiver should be able to pick out the signal that is meant for it. If different phases of the same spreading code is used, then system synchronization must be kept very precise to keep the different signals from overlapping. These and other related problems are very interesting by themselves, and how well they are overcome during system implementation can affect greatly the actual behavior of the system. Here we shall ignore them by assuming that the system we are dealing with is ideal. By doing so, we can isolate the effects on performance caused by the multiple-access protocol alone and study them on their own. Performance of actual systems can then be compared to this ideal system to see how good the implementation is.

Manuscript received May 7, 1995; revised October 10, 1995. This work was supported by the Hong Kong Research Grant Council and by the CRCG of the University of Hong Kong, Hong Kong. This paper was presented in part at the 46th IEEE Vehicular Technology Conference, Atlanta, GA, April 28–May 1, 1996.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, (email: flo@hkueee.hku.hk; tsng@hkueee.hku.hk; tiyuk@hkueee.hku.hk).

Publisher Item Identifier S 0733-8716(96)05236-5.

The model for the network is presented in Section II, and this simple model is analyzed in Section III using a discrete Markov chain. The throughput and delay results are presented in Section IV. In Section V, an approximate method for analyzing systems with a large number of users called equilibrium point analysis (EPA) is described and then used to analyze the proposed system. This section also shows how the stability of the system can be determined. Some numerical results are presented in Section VI. To demonstrate that the added complexity required by the proposed network does improve performance dramatically over simple CSMA-CD networks, a simple CSMA-CD network is analyzed using EPA in Section VII. A comparison of the delay-throughput characteristics of the two networks is then presented in Section VIII and the paper ends with some conclusions in Section IX.

II. Network Model

Time is divided into minislots and each station is assigned a different spreading sequence [6] or a different chip phase of the same maximal length spreading sequence [7], [8], if the stations are synchronized, with which it will receive messages. Thus the bandwidth of the system is divided into "virtual channels," one for each station. Messages are generated at each of the stations at a similar rate of s messages per minislot independent of other stations and other messages, and each station is equally likely to transmit to any other stations. We shall assume that $s \leq 1$, as the stations in our system have only one buffer slot to hold messages. Hence, s can also be interpreted as the message generation probability of each station per minislot. This assumption does not pose a big restriction to our analysis since ALOHA access works well only when $s \ll 1$. The messages are assumed to have a random number of minipackets, each one of a minislot in length, which are geometrically distributed with an average of l.

If a station has no message to transmit, it is said to belong to the idle mode T_0 and has an empty buffer to hold a message. Whenever a message arrives at an idle station during a minislot for station k, the station will listen to the channel of the intended receiver. If the channel is quiet, the station will attempt with probability one to transmit the packet during the next minislot. If no other packets are transmitted to the same receiving station during that same minislot, the station will capture the channel and continue transmitting until the whole message is successfully transmitted. If a collision occurs or if the channel is sensed busy, then the station will enter the blocked mode R_k. Blocked stations in R_k will not accept new messages that are generated, and will listen to the channel k during all the following minislots. When the channel is sensed free, the blocked station will attempt to transmit with probability p in the following minislot. Since the average length of the messages is l, once a station starts transmitting a message, the probability that the message will be completely sent during each subsequent minislot is $p_t = 1/l$.

As already stated, all timing imperfections, multipath fading, noise, and other sources of interference (such as finite cross-correlation between spread signals meant for different stations) are neglected in this model, and messages sent that do not collide are assumed to be received accurately. The minislots are assumed to be long enough so that acquisition of the spreading sequence can be achieved, collisions of messages can be detected by all stations, and the channels suffering from collisions can be left quiet before the start of the next minislot.

III. Analysis of Model

The number of idle stations that belong to the idle mode T_0 is denoted by n_0 and the number of blocked stations belonging to the blocked mode R_k is denoted by n_k. The state of the system during a minislot can then be denoted by the state vector $(n_1 t_1 n_2 t_2 \cdots n_N t_N)$ where t_k is 0 or 1, and is represented by a blank or t, depending on whether there is a station which has captured channel k during that minislot. Note that $\sum_{\text{all } i}(n_i + t_i) = N$—the total number of stations.

All rearrangements of $n_k t_k$ in $(n_1 t_1 n_2 t_2 \cdots n_N t_N)$ will be called substates of the same state and are equally likely since all the stations are statistically identical. For example, $(0\ 0t\ 1)$, $(0\ 1\ 0t)$, and $(1\ 0\ 0t)$ are all of the same state in a three-user system. They all show that one station is blocked trying to transmit to another while another station has successfully captured a channel. Note that each substate may have several configurations. For example, $(0\ 0t\ 1)$ can mean either station 1 is successfully transmitting to station 2, with station 2 blocked trying to transmit to station 3, or station 3 successfully transmitting to station 2 and station 1 blocked trying to transmit to station 3. Even though any of the substates can be used to represent that state, we shall use the substate where $n_i + t_i > n_j + t_j$ if $i < j$, and where $n_i + t_i = n_j + t_j$, then $t_i > t_j$ if $i < j$, to represent the state. We shall also drop the 0 for $0t$ for simplicity. Thus $(t\ 1\ 0)$ will be used to represent the above three substates. The states for systems with any number of users can then be written down easily by listing them systematically, as is demonstrated below for a system with three stations

$$(0\ 0\ 0)$$
$$(1\ 0\ 0)$$
$$(t\ 0\ 0)$$
$$(2\ 0\ 0)$$
$$(1t\ 0\ 0)$$
$$(1\ 1\ 0)$$
$$(t\ 1\ 0)$$
$$(t\ t\ 0)$$
$$(2\ 1\ 0)$$
$$(2\ t\ 0)$$
$$(1t\ 1\ 0)$$
$$(1t\ t\ 0)$$
$$(t\ 1\ 1)$$
$$(t\ t\ 1)$$
$$(t\ t\ t)$$

As an example, the state $(2\ t\ 0)$ means that two stations are blocked trying to communicate to a particular station, and one is successfully transmitting to another. Note also that $(N\ 0 \cdots 0), ((N-1)t0 \cdots 0)$, and $(1\ 1 \cdots 1)$ are not valid states.

Since the state a system is in at any particular minislot time depends only on the state the system was in at the previous slot, the evolution of the system states form a Markov process. Furthermore, since the state space is finite, equilibrium or stationary probabilities exist for all the states. If we write π_i as the stationary probability that the system is in state i, then we can, as shown in Appendix A, write the system equations in matrix form as

$$\begin{bmatrix} p_{11} & p_{21} & p_{31} & \cdots & p_{n1} \\ p_{12} & p_{22} & \cdots & & p_{n2} \\ p_{13} & \cdots & & & \\ \vdots & & & & \vdots \\ p_{1n} & p_{2n} & \cdots & & p_{nn} \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \vdots \\ \pi_n \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \vdots \\ \pi_n \end{bmatrix} \tag{1}$$

where p_{ij} is the sum of probabilities of any particular substate in state i changing to all substates in j.

To find p_{ij}, we can take a particular substate of state i, say $(i_1 t_{i1} i_2 t_{i2} \cdots i_N t_{iN})$, and find the probability that it will change to a substate of j, say $(j_1 t_{j1} j_2 t_{j2} \cdots j_N t_{jn})$. This probability can be found by using

$$\begin{aligned} p(i_1 \to j_1, t_{i1} \to t_{j1}, i_2 \to j_2, t_{i2} \to t_{j2}, \cdots,\\ i_N \to j_N, t_{iN} \to t_{jN}) \\ = p(i_1 \to j_1, t_{i1} \to t_{j1}) \\ \times p(i_2 \to j_2, t_{i2} \to t_{j2} \mid i_1 \to j_1 t_{i1} \to t_{ji}) \times \cdots \\ \times p(i_N \to j_N, t_{iN} \to t_{jN} \mid i_1 \to j_1, t_{i1} \to t_{j1}, \cdots,\\ i_{N-1} \to j_{N-1}, t_{iN-1} \to t_{jN-1}). \end{aligned}$$

If we write $p_k(i,j) = p(i_k \to j_k, t_{ik} \to t_{jk}|$ all modes $R_m, m < k$ already considered), we obtain (2), see bottom of the page, where $P_k(m \text{ mess})$ is the probability that m messages are generated during the minislot destined for station k. Thus

$$\begin{aligned} P_k(m \text{ mess}) = P(\text{station } k \text{ not in } T_0) \binom{r_k}{m} \left(\frac{s}{N-1}\right)^m \\ + P(\text{station } k \text{ in } T_0) \binom{r_k - 1}{m} \left(\frac{s}{N-1}\right)^m. \end{aligned} \tag{3}$$

Note that (3) is valid only for $m \geq 1$. The term r_k is the remaining number of idle users, or stations, at the time R_k is considered. The expression for $P(\text{station } k \text{ not in } T_0)$ and $P(\text{station } k \text{ in } T_0)$ are derived in Appendix B. Note that $p_N(i,j)$ involves the extra factor $(1-s)^{r_N+1}$ which is the probability that all unaffected idle stations do not transmit at all. Also $r_{k+1} = (r_k -$ number of newly generated messages), and $p_k(i,j) = 0$ if $r_{k+1} < 0$. Then $p_{ij} = \sum_{\text{all substates} \in j} \prod_{k=1}^{N} p_k(i,j)$.

IV. Results of Markov Analysis

By using the system of (1), together with the normalizing condition $\sum_{i=1}^{n} \pi_i = 1$, we can find the stationary probabilities of all the states $(\pi_1, \pi_2, \cdots, \pi_n)$. Suppose the state i has n_{ti} successfully transmitting stations, i.e., $n_{ti} = \sum_{i=1}^{N} t_i$. Then the probability that a message will be completely transmitted by the end of the minislot is just $n_{ti} p_t$. Thus the total average message throughput of the system is

$$S_{\text{total}} = \sum_{i=1}^{n} \pi_i n_{ti} p_t. \tag{4}$$

The total average number of blocked stations for state i is $\pi_i \sum_{k=1}^{N} (n_k \text{ for state } i)$. The total average number of blocked stations for all states is, therefore

$$b = \sum_{i=1}^{n} \pi_i \sum_{\text{all } k} (n_k \text{ for state } i). \tag{5}$$

By using Little's Theorem [9], the total delay suffered by each message is found to be

$$D = \frac{b}{S_{\text{total}}}. \tag{6}$$

Fig. 1 shows the throughput delay curves with $s \cdot l$ varying from 0.1 to 0.9 for four different values of p, for a fully-connected CDMA ALOHA network of five stations with channel sensing and collision detection, together with some simulation points. The average message length is ten minislots,

$$p_k(i,j) = \begin{cases} 0, & j_k - i_k < -1 \\ 0, & j_k - i_k = -1 \text{ and } t_{jk} - t_{ik} \neq 1 \\ i_k p(1-p)^{i_k - 1}, & j_k - i_k = -1 \text{ and } t_{jk} - t_{ik} = 1 \\ 1 - i_k p(1-p)^{i_k - 1}, & j_k - i_k = 0 \text{ and } t_{jk} = t_{ik} = 0 \\ (1 - p_t), & j_k - i_k = 0 \text{ and } t_{jk} = i_{jk} = 1 \\ (1-p)^{i_k} P_k(1 \text{ mess}), & j_k - i_k = 0 \text{ and } t_{jk} - t_{ik} = 1 \\ p_t, & j_k - i_k = 0 \text{ and } t_{jk} - t_{ik} = -1 \\ (1 - (1-p)^{i_k}) P_k(1 \text{ mess}), & j_k - i_k = 1 \text{ and } t_{jk} = t_{ik} = 0 \\ (1 - p_t) P_k(1 \text{ mess}), & j_k - i_k = 1 \text{ and } t_{jk} = t_{ik} = 1 \\ 0, & j_k - i_k = 1 \text{ and } t_{jk} - t_{ik} = 1 \\ p_t P_k(1 \text{ mess}), & j_k - i_k = 1 \text{ and } t_{jk} - t_{ik} = -1 \\ P_k(m \text{ mess}), & j_k - i_k = m > 1 \text{ and } t_{jk} = t_{ik} = 0 \\ (1 - p_t) P_k(m \text{ mess}), & j_k - i_k = m > 1 \text{ and } t_{jk} = t_{ik} = 1 \\ 0, & j_k - i_k = m > 1 \text{ and } t_{jk} - t_{ik} = 1 \\ p_t P_k(m \text{ mess}), & j_k - i_k = m > 1 \text{ and } t_{jk} - t_{ik} = -1 \end{cases} \tag{2}$$

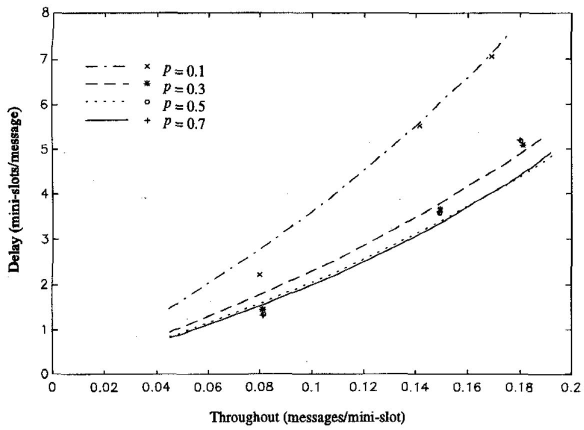


Fig. 1. Throughput-delay curves for a system with five users using CDMA ALOHA with channel sensing and collision detection. The simulation points are for $s \cdot l = 0.2, s \cdot l = 0.5$, and $s \cdot l = 0.8$. The average message length is ten minislots. Simulations are over 200 000 minislots.

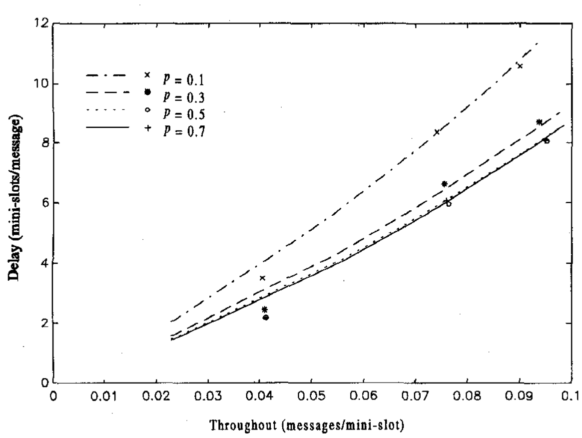


Fig. 2. Throughput-delay curves for a system similar to that of Fig. 1, but now with average message lengths of 20 minislots. Simulation points are for $s \cdot l = 0.2, s \cdot l = 0.5$, and $s \cdot l = 0.8$. Simulations are done over 200 000 minislots.

and the simulations are done over 200 000 minislots. We use $s \cdot l$, the message generation probability per average message, rather than s, the message generation probability per minislot. This is to avoid situations where many newly generated messages are dropped because older messages are not yet completely transmitted.

From the graph, it can be seen that throughput is more dependent on s, as the systems with different p have almost the same throughput for the same s. The delay depends on p. With a small p, delay is high because a lot of minislots are unused. As p increases, delay drops, until about $p = 0.5$. After that delays caused by collisions between retransmitted messages make up for the lack of delay due to unused minislots.

Fig. 2 shows the throughput-delay curves for a similar system, but now with message lengths of 20 minislots. The qualitative behavior is very close to that of the previous system. Throughput is almost halved and delay is almost doubled, although the delay performance seems proportionately better than that of the system where message sizes are halved.

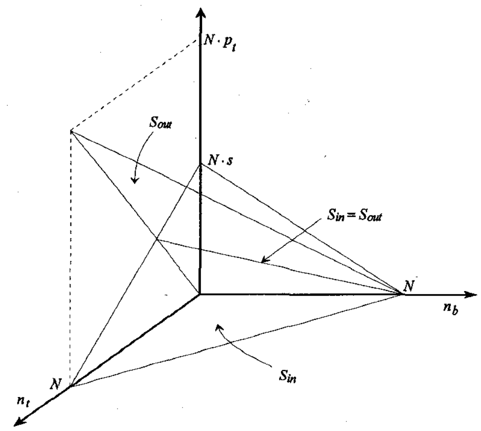


Fig. 3. Graph showing the $S_{\rm in}$ and $S_{\rm out}$ planes, and their intersection.

For both message lengths, a system with five stations does not experience excessive delays. Thus these systems are stable, at least for the values of s and p studied.

V. Equilibrium Point Analysis of Multichannel Networks

The analysis proposed can be used, in principle, to obtain the stationary probabilities for a system of any size. The amount of computation, however, rises rapidly with the number of users. For the system with just five stations, the number of states is already 71 and the number of substates, 1672.

A much simpler approach that can be used for large systems is EPA, first introduced in [10] for simple slotted ALOHA. The idea behind EPA is that in the long run, the system will stay near the states where the number of messages generated will balance the number of messages successfully transmitted. In our case, since a station must first capture the channel before it can start successful transmission, the capture rate too must match the message generation and message completion rates.

The message generation rate only depends on n_0, the number of idle stations during a minislot, and it is

$$S_{\rm in} = n_0 s = (N - n_t - n_b)s \tag{7}$$

where n_b is the total number of blocked stations and n_t is the total number of transmitting stations. The message completion rate is only dependent on n_t and it is

$$S_{\rm out} = n_t p_t. \tag{8}$$

When plotted against n_b and n_t, $S_{\rm in}$ and $S_{\rm out}$ form planes as shown in Fig. 3, and their intersection is the line where

$$n_t = \frac{s(N - n_b)}{s + p_t}, \qquad 0 \le n_b \le N. \tag{9}$$

At equilibrium, $S_{\rm in}$ is equal to $S_{\rm out}$, and both must be equal to the rate of capture of channels, $S_{\rm cap}$. By obtaining this point, we can find the equilibrium throughput and also the number of blocked stations at equilibrium. These equilibrium values we shall then assume to be the averages of the two quantities.

We shall consider a network of N stations which behave as described in Section II. For ease of discussion, we shall call channels which have blocked stations "occupied." First, we wish to determine if a system will be stable and uncongested, as defined in [10]. Consider the situation where an occupied channel k has n_k blocked stations, and there are n_0 idle stations in the system. When the channel k is captured, on average, per minislot, the number of idle stations that will be presented with a newly generated message intended for station k is

$$\begin{aligned}\Delta n_k(\text{cap}) = {} & n_0\frac{s}{N}\left(1-\frac{s}{N}\right)^{n_0-1} \\ & + 2\binom{n_0}{2}\left(\frac{s}{N}\right)^2\left(1-\frac{s}{N}\right)^{n_0-2} \\ & + 3\binom{n_0}{3}\left(\frac{s}{N}\right)^3\left(1-\frac{s}{N}\right)^{n_0-3} + \cdots \\ & + n_0\left(\frac{s}{N}\right)^{n_0}. \end{aligned} \tag{10}$$

This is also, of course, the average number of additional blocked stations that will appear at the channel k. Note that s/N is an approximation of $s/(N-1)$, and that we assume that station k is not idle. The error introduced by this assumption is significant only when n_0 is small and, as will be seen later, only affect judgment on systems that are marginally unstable.

When the channel is free, on average per minislot, the additional number of blocked stations that will appear at the channel k is

$$\begin{aligned}\Delta n_k(\text{free}+) = {} & n_0\frac{s}{N}\left(1-\frac{s}{N}\right)^{n_0-1}[1-(1-p)^{n_k}] \\ & + 2\binom{n_0}{2}\left(\frac{s}{N}\right)^2\left(1-\frac{s}{N}\right)^{n_0-2} \\ & + 3\binom{n_0}{3}\left(\frac{s}{N}\right)^3\left(1-\frac{s}{N}\right)^{n_0-3} + \cdots \\ & + n_0\left(\frac{s}{N}\right)^{n_0}. \end{aligned} \tag{11}$$

When the channel is free, on average per minislot, the decrease in the number of blocked stations is just the probability that a blocked station will capture the channel. This is

$$\Delta n_k(\text{free}-) = n_k p(1-p)^{n_k-1}\left(1-\frac{s}{N}\right)^{n_0}. \tag{12}$$

On average, a channel will remain captured for $l+1$ minislots, as it takes l minislots on average to complete transmission, and an extra minislot for all the stations to realize that the channel is now free, and it will remain free for l_f minislots, with $l_f = 1/p_f$, where p_f is the probability that a station will capture the channel

$$\begin{aligned}p_f = {} & n_0\frac{s}{N}\left(1-\frac{s}{N}\right)^{n_0-1}(1-p)^{n_k} \\ & + n_k p(1-p)^{n_k-1}\left(1-\frac{s}{N}\right)n_0. \end{aligned} \tag{13}$$

Thus, on average, the number of additional blocked stations that will appear at the blocked channel k per minislot is just

$$\begin{aligned}\Delta n_k = {} & \frac{l+1}{l+1+l_f}[\Delta n_k(\text{cap})] + \frac{l_f}{l+1+l_f} \\ & \times [\Delta n_k(\text{free}+) - \Delta n_k(\text{free}-)]. \end{aligned} \tag{14}$$

For the system to be at equilibrium, the average message input S_{in} must be equal to the average message completion S_{out}, i.e., (9) must be satisfied. To determine whether a system at equilibrium is stable, we examine all the points where (9) is satisfied, and calculate Δn_k given by (14) for $n_k = \{1, 2, 3, \cdots, n_b\}$, and for $0 < n_b < N$. If all the Δn_k are negative, then the system is stable since any occupied channel will tend to clear itself of blocked stations. For cases where all Δn_k are positive, then the system is congested, as all occupied channels tend to accumulate blocked stations, and thus lower the message input. For cases where Δn_k can be positive or negative, the system is unstable, since it can remain at points in phase space where message input is high or at points where most of the stations are blocked.

Once we have determined that a system is stable and uncongested, we have to estimate its throughput and delay. If a system is in equilibrium, as we have seen, the number of transmitting stations and blocked stations must be related by (9). We have to find out the point on the equilibrium curve determined by this equation where the system will settle down. To do this, we have to calculate the equilibrium capture rate. The capture rate of the system given n_t and n_b depends on how the blocked stations are distributed among the occupied channels. We have already seen that for the system to be uncongested and stable, the occupied channels will tend to clear themselves. Thus, it is reasonable to assume that at equilibrium, stable and uncongested networks will have only one blocked station per occupied channel. The results presented in the next section shows that this assumption is valid.

Suppose there are n_b blocked stations distributed one to each occupied channel. For occupied channels, the capture rate when the channel is free is

$$S_{\text{cap}}(\text{occ}) = n_0\frac{s}{N}\left(1-\frac{s}{N}\right)^{n_0-1}(1-p) + \left(1-\frac{s}{N}\right)^{n_0}p. \tag{15}$$

On average, the occupied channel will stay free for $l_c = 1/S_{\text{cap}}(\text{occ})$ minislots. For an unoccupied channel, the capture rate when the channel is free is just

$$S_{\text{cap}}(\text{unocc}) = n_0\frac{s}{N}\left(1-\frac{s}{N}\right)^{n_0-1}. \tag{16}$$

On average, the unoccupied channel will stay free for $l_u = 1/S_{\text{cap}}(\text{unocc})$ minislots. Note that in the derivation of (15) and (16), we have implicitly assumed that the station which owns the channel is not among the idle stations. The error introduced by this assumption is small as long as n_0 is large, and this is true for all uncongested systems.

When a channel is captured, its capture rate of course drops to zero. Furthermore, a channel will stay captured for $l+1$ minislots on average. Thus the average capture rate of a stable, uncongested system at equilibrium is

$$\begin{aligned}S_{\text{cap}}(\text{average}) = {} & n_b\frac{l_c}{l_c+l+1}S_{\text{cap}}(\text{occ}) \\ & + (N-n_b)\frac{l_u}{l_u+l+1}S_{\text{cap}}(\text{unocc}) \\ = {} & n_b\frac{1}{l_c+l+1} + (N-n_b)\frac{1}{l_u+l+1}. \end{aligned} \tag{17}$$

When we plot $S_{\text{cap}}(\text{average})$ and the $S_{\text{in}} = S_{\text{out}}$ curve against n_b and n_t related by (9), we will get just one intersection. For

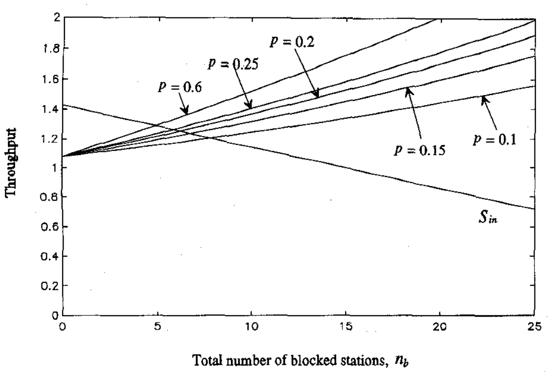


Fig. 4. Graph of the message input and the average message output curves for $N = 50$, $s = 0.04$, and $l = 10$, for various values of p for a multichannel system.

TABLE I
THROUGHPUTS AND DELAYS FOR A MULTICHANNEL SYSTEM WITH $N = 50, s = 0.004, l = 10$ AND VARIOUS VALUES OF p. SIMULATIONS ARE DONE OVER 100 000 MINISLOTS

				EPA		Simulation	
s	p	ℓ	S_{cap}	delay	status	thruput	delay
0.04	0.10	10	1.19	6.78	stable	1.1783	6.81
0.04	0.15	10	1.23	5.71	stable	1.2009	6.07
0.04	0.20	10	1.24	5.15	unstable	1.2170	5.52
0.04	0.25	10	1.25	4.80	unstable	1.2295	5.25
0.04	0.60	10	1.28	3.97	unstable	0.0500	965.02

stable and uncongested system, this point gives the equilibrium throughput S_{out} and number of blocked stations n_b. Using Little's Theorem, we can find the equilibrium delay, $d = n_b/S_{\text{out}}$.

VI. RESULTS FOR MULTICHANNEL SYSTEMS

Fig. 4 shows the S_{in} and the $S_{\text{cap}}(\text{average})$ curves for $s = 0.04, l = 20$ and the various values of p indicated. When the system is determined to be stable, the intersection point of the two curves give the equilibrium throughput and number of blocked stations for the system.

Table I shows the throughput and delay results obtained from EPA and simulations over 100 000 minislots of a multichannel CSMA-CD system with $N = 50, l = 10$, and $s = 0.04$. For $p = 0.10$ and $p = 0.15$, the system is determined to be stable as all Δn_k are negative. For $P = 0.20$ and $p = 0.25$, the system is unstable, but the system will tend toward the congested region only when n_k becomes large. For the $p = 0.20$ case, n_k must cross 27 and for $p = 0.25$ case, n_k must cross 20 before Δn_k becomes positive. The system will therefore stay in the uncongested region for a long time before drifting into the congested part of the system phase space. The throughputs and delays obtained from simulations show that 100 000 minislots are not sufficient for the system to become congested.

For the $p = 0.60$ case, n_k needs only to be greater than 6 before Δn_k becomes positive. Thus it is relatively easy for the system to move to the congested part of the system phase space. Simulation over 100 000 slots shows indeed a very low throughput and high delay.

VII. EPA OF SINGLE CHANNEL NETWORKS

To show that CDMA is worth employing in our proposed network, we next show how EPA can be used to analyze a single channel CSMA-CD network, and later compare the performance of this network with that of the multichannel one.

Let us consider a single channel minislotted CSMA-CD network with N stations. It is similar to that described in Section II, except that CDMA is not used and therefore all stations share a single transmission channel. Assume now that the system is in equilibrium with n_{be} blocked stations. The number of idle stations is $n_{oe} = N - n_{be}$. The rate at which new messages are generated is $S_{\text{in}} = n_{oe}S$. When the channel is captured, no stations can capture the channel. When the channel is free, the capture rate is

$$S_{\text{cap}}(\text{free}) = n_{oe}s(1-s)^{n_{oe}-1}(1-p)^{n_{be}} + n_{be}p(1-p)^{n_{be}-1}(1-s)^{n_{oe}}. \quad (18)$$

The channel will remain captured on average $l + 1$ minislots, and the channel will remain free on average $l_f = 1/S_{\text{cap}}(\text{free})$ minislots. The average capture rate of the channel is therefore

$$S_{\text{cap}}(\text{average}) = \frac{l_f}{l+1+l_f} S_{\text{cap}}(\text{free}) = \frac{l}{l+1+l_f}. \quad (19)$$

Since any station capturing the channel is guaranteed to successfully complete transmitting its message, therefore the average capture rate is also the average message output rate S_{out}.

Thus by plotting S_{in} and S_{cap} (average) against n_b, we can determine the equilibrium point(s) by noting the intersection(s) of the two curves. Following [10], we can have the following situations. If there is only one intersection point where n_b is much less than N, then the system is stable with high throughput. If more than one intersection point exist, then the system is not stable as the system will move from one stable equilibrium point to another with the passage of time. If only one equilibrium point exists at n_b neaer N, then the system is said to be congested, with very low throughput and very high delay. The equilibrium delay, d, can be obtained from S_{out} and n_b by Little's Formula, $d = n_b/S_{\text{out}}$.

Note that equilibrium points with $S_{\text{in}} < S_{\text{out}}$ to the left and $S_{\text{in}} > S_{\text{out}}$ to the right are stable equilibrium points since any movement away from them will cause the system to shift back toward those points. The other type of equilibrium point is unstable.

Fig. 5 shows the S_{in} and S_{out} curves of a single-channel CSMA-CD system with $N = 50, s = 0.001$, and $l = 20$. When $p = 0.1$, the system is stable with high throughput. When $p = 0.15$, the system is unstable, but there is a big separation between the first two equilibrium points, meaning that the system will require a long time to move from one to the other. When $p = 0.2$, the separation between the first two equilibrium points is much smaller. When $p = 0.22$, the system is congested and throughput drops to a very low number.

Table II shows the throughputs and delays of the system obtained by EPA and simulation for 100 000 minislots for the various values of p plotted in Fig. 5. When the system is stable

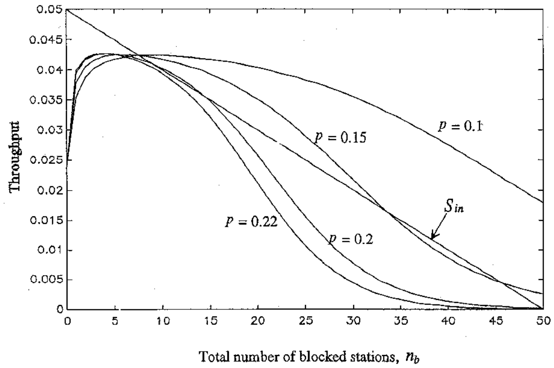


Fig. 5. Graph of the message input and average message output curves for Equation $N = 50, s = 0.001, l = 20$ and various values of p, for a single-channel system.

TABLE II

THROUGHPUTS AND DELAYS FOR A SINGLE-CHANNEL SYSTEM WITH $N = 50, s = 0.001, l = 20$ AND VARIOUS VALUES OF p. NOTE THAT SIMULATIONS ARE DONE OVER 100 000 MINISLOTS, AND FOR UNSTABLE SYSTEMS, THE EQUILIBRIUM VALUES FOR THE FIRST EQUILIBRIUM POINT ARE TABULATED

s	p	ℓ	S_{cap}	EPA delay	status	Simulation thruput	Simulation delay
10^{-3}	0.10	20	0.0423	181.1	stable	0.0429	135.5
10^{-3}	0.15	20	0.0424	178.3	unstable	0.0421	156.1
10^{-3}	0.20	20	0.0410	218.3	unstable	0.0106	3639.0
10^{-3}	0.22	20	0.0001	498600	congested	0.0019	25502

TABLE III

THROUGHPUTS AND DELAYS FOR A MULTICHANNEL SYSTEM WITH $N = 50$, AND VARIOUS VALUES OF s, p, AND l. NOTE THAT SIMULATIONS ARE DONE OVER 100 000 MINISLOTS

s	p	ℓ	S_{cap}	EPA delay	status	Simulation thruput	Simulation delay
10^{-3}	0.05	10	0.0495	0.40	stable	0.050	0.01
10^{-3}	0.05	20	0.0490	1.02	stable	0.050	0.15
10^{-3}	0.10	10	0.0495	0.20	stable	0.050	0.00
10^{-3}	0.10	20	0.0490	0.61	stable	0.050	0.15
2×10^{-3}	0.05	10	0.0979	0.71	stable	0.100	0.01
2×10^{-3}	0.05	20	0.0958	1.77	stable	0.100	0.01
2×10^{-3}	0.10	10	0.0979	0.51	stable	0.100	0.01
2×10^{-3}	0.10	20	0.0959	1.35	stable	0.098	0.01
10^{-3}	0.20	20	0.0490	0.61	stable	0.050	0.00
2×10^{-3}	0.20	20	0.0959	1.15	stable	0.098	0.01

TABLE IV

THROUGHPUTS AND DELAYS FOR A SINGLE-CHANNEL SYSTEM WITH $N = 50$, AND THE SAME VALUES OF s, p, AND l AS IN TABLE III. NOTE THAT SIMULATIONS ARE DONE OVER 100 000 MINISLOTS, AND FOR THE UNSTABLE SYSTEM, THE EQUILIBRIUM VALUES FOR THE FIRST EQUILIBRIUM POINT ARE USED

s	p	ℓ	S_{cap}	EPA delay	status	Simulation thruput	Simulation delay
10^{-3}	0.05	10	0.0487	26.9	stable	0.0495	17.4
10^{-3}	0.05	20	0.0412	213.6	stable	0.0418	156.7
10^{-3}	0.10	10	0.0494	13.2	stable	0.0496	15.8
10^{-3}	0.10	20	0.0423	181.1	stable	0.0423	148.0
2×10^{-3}	0.05	10	0.0728	186.5	stable	0.0718	159.1
2×10^{-3}	0.05	20	0.0417	697.8	stable	0.0420	626.0
2×10^{-3}	0.10	10	0.0720	194.2	stable	0.0716	167.3
2×10^{-3}	0.10	20	0.0329	1019.2	stable	0.0335	933.6
10^{-3}	0.20	20	0.0411	215.3	unstable	0.0057	7712.5
2×10^{-3}	0.20	20	0.0002	249400	congested	0.0008	62512.9

with $p = 0.10$, the throughput and delay values obtained from EPA match those obtained from simulation, as expected. For $p = 0.15$ simulation over 100 000 minislots is insufficient to allow the system to reach the second stable equilibrium point, and system performance matches that obtained from the first equilibrium point. For the $p = 0.20$ case, the system reaches the second stable equilibrium point easily, and system performance drops drastically. For the congested $p = 0.22$ case, both EPA and simulation provide very low throughput and very high delay values.

VIII. COMPARISON OF SINGLE AND MULTICHANNEL SYSTEMS

Tables III and IV show the throughput-delay figures of the multi as well as the single-channel CSMA-CD networks obtained from both EPA, as well as simulations over 100 000 minislots, for the same set of the parameters s, p, and l. Looking at Tables III and IV, we can see that when the system is lightly loaded, e.g., with $s = 0.001, l = 10$, and $p = 0.005$ and 0.10, both the multichannel as well as single-channel CSMA-CD networks can handle almost the maximum message generation rate, $Ns = 0.05$. The delays suffered by the multichannel system, however, are much lower.

The difference in performance is even more evident when the system is more heavily loaded. For $s = 0.002, l = 20$, and $p = 0.1$, the single channel system can handle only about 33% of the maximum possible number of messages generated, with delays suffered by successful messages going beyond 1000 minislots. The multichannel system, on the other hand, can still handle about 96% of the maximum possible number of messages generated, with successful messages suffering minimal delays. Even when the single channel system enter the unstable and congested region with $p = 0.20$, the multichannel system can still handle above 95% of the maximum number of messages generated, with negligible message delays.

IX. CONCLUSION

We have presented a method of determining the exact average throughput and delay performance of a CDMA ALOHA network with channel sensing and collision detection using Markov analysis. This method is suitable for networks with a small number of stations. We have also shown that for networks with a large number of stations, EPA is a more appropriate analysis tool. It allows us to determine if a system is stable, and if it is so, to estimate the equilibrium throughput and delay. Using EPA, we have shown that messages transmitted in CDMA ALOHA networks using channel sensing and collision detection with a large number of stations have very low delays. The analyses provided in this paper also allow a designer of such networks to keep the delay at a minimum, by choosing the highest possible retransmission probability while still keeping the system stable.

We have also shown that CSMA-CD networks that use CDMA to create "virtual channels" have much better delay-throughput characteristics than simple single-channel CSMA-CD systems. One final point we wish to stress is that multichannel CDMA CSMA-CD systems do not only provide higher throughputs and lower delays than single-channel systems. Even though they require slightly more complex transmission and reception equipment, they also pose, when lightly loaded, almost no interference to other existing narrow-band users.

APPENDIX A
SIMPLIFICATION OF TRANSITION MATRIX

If we write π_{ij} as the stationary probability that the system is in state i substate j, or in shorthand form i, j, and P_{ij}, k_l as the probability that $\{i,j\}$ will change to $\{k,l\}$ in a slot time, then we can write the system equations in matrix form as (A.1), see bottom of the page, where $k_1, k_2, \cdots, k_n$ are the numbers of substates of state $1, 2, \cdots, n$, respectively. Now $\pi_{11} = \pi_{12} = \cdots = \pi_{1k_1}$ and $\pi_{21} = \pi_{22} = \cdots = \pi_{2k_2}$, etc., since all substates of a certain state are equally probable. Let us denote the difference of blocked stations and transmitting stations in the various modes between $\{i,j\}$ and $\{k,l\}$ by $(d_1t_1d_2t_2\cdots d_Nt_N)$.. A little thought will show that these differences between a certain substate, say one, of i and all substates of k, and another substate of i, say two, and all substates of k, are just rearrangements of the same set of $(d_1t_1d_2t_2\cdots d_Nt_N)$. Since the probability that a certain $\{i,j\}$ will change to a certain $\{k,l\}$ after a slot depends only on the difference in the number of blocked and transmitting stations in the various modes, hence $(p_{11,11}+p_{11,12}+\cdots+p_{11,1k_1}) = (p_{12,11}+p_{12,12}+\cdots+p_{12,1k_1}) = \cdots = (p_{1k_1,11}+p_{1k_1,12}+\cdots+p_{1k_1,1k_1}) \equiv p_{11}$ and $(p_{21,11}+p_{21,12}+\cdots+p_{21,1k_1}) = (p_{22,11}+p_{22,12}+\cdots+p_{22,1k_1}) = \cdots = (p_{2k_2,11}+p_{2k_2,12}+\cdots+p_{2k_2,1k_1} \equiv p_{21}$, etc. Multiplying out the first k_1 rows of (A.1) and adding, we get

$$p_{11}k_1\pi_{11} + p_{21}k_2\pi_{21} + \cdots + p_{n1}k_n\pi_{n1} = k_1\pi_{11}. \tag{A.2}$$

Let $k_1\pi_{11} = \pi_1, k_2\pi_{21} = \pi_2, \cdots, k_n\pi_{n1} = \pi_n$. Note that π_i is the sum of probabilities of substates in state i, and is therefore the probability of state i. Thus, we get $p_{11}\pi_1 + p_{21}\pi_2 + \cdots + p_{n1}\pi_n = \pi_1$. Similarly, multiplying out the next k_2 rows of (A.1) and adding, and so forth, we can get $p_{12}\pi_1 + P_{22}\pi_2 + \cdots + p_{n2}\pi_n = \pi_2, \cdots, p_{1n}\pi_1 + p_{2n}\pi_2 + \cdots + p_{nn}\pi_n = \pi_n$. The system of equations is then simplified to

$$\begin{bmatrix} p_{11} & p_{21} & p_{31} & \cdots & p_{nl} \\ p_{12} & p_{22} & \cdots & & p_{n2} \\ p_{13} & \cdots & & & \\ \vdots & & & & \vdots \\ p_{1n} & p_{2n} & \cdots & & p_{nn} \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \vdots \\ \pi_n \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \vdots \\ \pi_n \end{bmatrix} \tag{A.3}$$

where p_{ij} is the sum of probabilities of any particular substate in state i changing to all substates in j, and π_i is the stationary probability of the system in state i.

APPENDIX B
DERIVATION OF P(station k in T_0)

Consider the substate $m = (n_1t_1n_2t_2\cdots n_Nt_N)$. If we lay out the users in the various modes in a row we get

$$\begin{array}{cccccc} R_1 & R_2 & R_3 & \cdots & R_N & T_0 \\ XX..X & X..X & X..X & & X..X & XX..X \\ m_1 & m_2 & m_3 & \cdots & m_N & m_0 \end{array}$$

where $m_i = n_i + t_i$.

The total number of sequences without restrictions is $N!$. But user 1 cannot be in R_1, so we need to subtract $m_1(N-1)!$ sequences. Similarly, user 2 cannot be in R_2, etc., so we need to subtract $(m_1 + m_2 + \cdots + m_N) \times (N-1)!$ sequences. Now we have oversubtracted as some of these sequences are the same, for example, those with users one and two in modes R_1 and R_2 respectively, so we need to add them back. The number of sequences with two particular users in their own modes is $(m_1m_2 + m_1m_3 + \cdots + m_2m_3 + m_2m_4 + \cdots + m_{N-1}m_N) \times (N-2)!$. Now we have overadded as some of these sequences have three stations in their own three modes. The number that do is $(m_1m_2m_3 + m_1m_3m_4 + \cdots + m_2m_3m_4 + \cdots + m_{N-2}m_{N-1}m_N) \times (N-3)!$. By carrying out this line of reasoning till the end, we find that the number of possible sequences is $N! - (\sum m_i)\colon (N-1)! + (\sum_{i\neq j} m_im_j)\cdot(N-2)! - \cdots + (-1)^N \cdot (m_1m_2\cdots m_N)$. To obtain the number of possible configurations, we need to weed out those sequences with the same stations in a mode. The number of sequences with m_i stations in R_1 that are the same is $m_1!$ since m_1 stations can be arranged $m_1!$ ways. This is true for m_2, etc. Thus, the number of possible configurations, β, is

$$\beta = \frac{N! - (\sum m_i)\cdot(N-1)! + \cdots + (-1)^N(m_1m_2\cdots m_N)}{m_1!m_2!\cdots m_N!(N - \sum m_1)!}. \tag{A.4}$$

To find the number of configurations with the station k in T_0, denoted by β_k, we first note that if all stations are idle, $\beta_k = 1$, and if all stations are blocked or transmitting, $\beta_k = 0$. If m stations, $0 < m < N$, are free, we can find β_k by first taking out station k and putting it in T_0. By using the same line of reasoning we have used before, and keeping in mind that we are now working with only $N-1$ stations since station k

$$\begin{bmatrix} p_{11,11} & p_{12,11} & \cdots & p_{1k_1,11} & p_{21,11} & \cdots & p_{2k_2,11} & \cdots & p_{nk_n,11} \\ p_{11,12} & p_{12,12} & \cdots & p_{1k_1,12} & p_{21,12} & \cdots & & & p_{nk_n,12} \\ \vdots & & & & & & & & \vdots \\ p_{11,1k_1} & P_{12,1k_1} & & & & & & & p_{nk_n,1k_1} \\ p_{11,21} & \cdots & & & & & & & \vdots \\ p_{11,22} & & & & & & & & \\ \vdots & & & & & & & & \\ p_{11,nk_{n-1}} & \cdots & & & & & & & \vdots \\ p_{11,nk_n} & \cdots & & & & & & \cdots & p_{nk_n,nk_n} \end{bmatrix} \times \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \vdots \\ \pi_{1k_1} \\ \pi_{21} \\ \vdots \\ \vdots \\ \pi_{nk_n} \end{bmatrix} = \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \vdots \\ \pi_{1k_1} \\ \pi_{21} \\ \vdots \\ \vdots \\ \pi_{nk_n} \end{bmatrix} \tag{A.1}$$

has been put in T_0, we get

$$\beta_k = \frac{(N-1)! - (\sum_{i \neq k} m_i) \cdot (N-2)! + \cdots + (-1)^N \prod_{i \neq k} m_i}{m_1! m_2! \cdots m_N! (N - \sum_{\text{all } i} m_i - 1)!}. \tag{A.5}$$

Then we get

$$P(\text{station } k \text{ in } T_0) = \beta_k / \beta. \tag{A.6}$$

References

[1] N. Abramson, "The ALOHA system—Another alternative for computer communications," in *Proc. AFIPS Conf., 1970 Fall Joint Comput. Conf.*, 1970, pp. 281–285.

[2] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part I—Carrier sense multiple access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. COM-23, pp. 1400–1416, 1975.

[3] F. Tobagi and V. Hunt, "Performance analysis of channel sense multiple access with collision detection," *Comput. Networks*, vol. 4, pp. 245–259, 1980.

[4] R. Metcalfe and D. Boggs, "Ethernet: Distributed packet switching for local computer networks," *Commun. ACM*, vol. 19, no. 7, pp. 395–404, 1976.

[5] R. Pickholtz, D. Schilling, and L. Milstein, "Theory of spread-spectrum communication—A tutorial," *IEEE Trans. Commun.*, vol. COM-30, pp. 855–884, 1982.

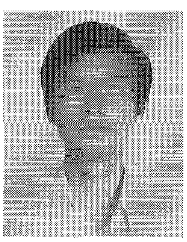
[6] E. Sousa and J. Silvester, "Spreading code protocols for distributed spread-spectrum packet radio networks," *IEEE Trans. Commun.*, vol. 36, pp. 272–281, 1988.

[7] N. Abramson, "Multiple access in wireless digital networks," *Proc. IEEE*, vol. 82, pp. 1360–1370, 1994.

[8] K. Yip and T. Ng, "Code phase assignment—A technique for high capacity indoor mobile DS-CDMA communication," in *IEEE 44th Veh. Technol. Conf.*, 1994, pp. 1586–1590.

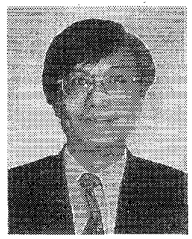
[9] J. Little, "A proof of the queuing formula $L = \lambda W$," *Oper. Res.*, vol. 9, pp. 383–387, 1961.

[10] L. Kleinrock and S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, vol. COM-23, pp. 410–423, 1975.



Fook Loong Lo (M'88) received the A.B. degree in physics from Hamilton College, Clinton, NY, in 1982, and the M.S.E.E. degree in communication theory and systems from the University of California, San Diego, in May 1985.

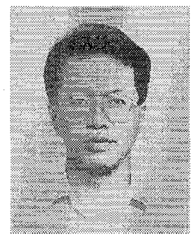
He taught at the Singapore Polytechnic, Singapore, from October 1985 to April, 1994. Since then, he has been with the Department of Electrical and Electronic Engineering, The University of Hong Kong, where he is a Ph.D. student and a Member of the Spread-Spectrum Communication Research Group. His technical interests include wireless computer networking, spread-spectrum communication, and network security.



Tung Sang Ng (S'74–M'78–SM'90) received the B.Sc.(Eng.) degree from The University of Hong Kong, in 1972, and the M.Eng.Sc. and Ph.D. degrees from the University of Newcastle, Australia, in 1974 and 1977, respectively, all in electrical engineering.

He was employed as an Electrical Engineer by BHP Newcastle Iron and Steel Works, Australia, from 1972 to 1974. From 1977 to 1990, he was a Lecturer, then Senior Lecturer and Reader in the Department of Electrical and Computer Engineering, The University of Wollongong, Australia. Currently, he is holding the Chair of Electronic Engineering at The University of Hong Kong. His current research interests include spread-spectrum techniques, digital signal processing, mobile communication systems, and engineering applications of artificial intelligence. He is a Regional Editor for the international journal *Engineering Applications of Artificial Intelligence*.

Dr. Ng is a Fellow of the IEE, HKIE, and IEAust.



Tony T. Yuk (S'82–M'84) received the B.S. degree from Iowa State University, IA, in 1978, and the M.S. and Ph.D. degrees from Arizona State University, AZ, in 1980 and 1986, respectively, all in electrical engineering.

Since 1986, he has been a Lecturer at The University of Hong Kong, Hong Kong. His current research interests are design and numerical modeling of fiber laser, bandwidth management, and protocol modeling of high speed networks and multimedia systems.